Applied Mathematical Sciences, Vol. 7, 2013, no. 24, 1209 - 1212
HIKARI Ltd, www.m-hikari.com

The Farthest Substring Problem

Anna Gorbenko

Department of Intelligent Systems and Robotics
Ural Federal University
620083 Ekaterinburg, Russia
gorbenko.ann@gmail.com

Vladimir Popov

Department of Intelligent Systems and Robotics
Ural Federal University
620083 Ekaterinburg, Russia
Vladimir.Popov@usu.ru

Abstract

In this paper we consider an approach to solve the farthest substring problem. This approach is based on an explicit reduction from the problem to the satisfiability problem.

Keywords: farthest substring problem, satisfiability, **NP**-complete

Usage of different regularities (see e.g. [1] – [3]) has become essential in modern computer science. We can mention algorithmic problems of bioinformatics (see e.g. [4] – [7]). Also, we can mention various problems of robotics, sensor placement (see e.g. [8, 9]), problems of selection of visual landmarks (see e.g. [10, 11]), technical vision (see e.g. [12] – [14]), robot self-awareness (see e.g. [15, 16]), planning (see e.g. [17] – [20]), etc. In this paper we consider the farthest substring problem. The problem is **NP**-complete for strings over any alphabet Σ with $|\Sigma| \geq 2$ [21].

Encoding hard problems as instances of SAT and solving them with efficient satisfiability algorithms has caused considerable interest (see e.g. [1] – [3]). In this paper, we consider an approach to solve the farthest substring problem. Our approach is based on an explicit reduction from the problem to the satisfiability problem. Let

$$\delta(X, S)$$

be the Hamming distance. The decision version of the farthest substring problem can be formulated as following.

The farthest substring problem (FSS):

Instance: *Given a set $\mathcal{M}$ of strings of length at least n over an alphabet Σ and a positive integer D.*

Question: *Is there a string X of length n over Σ such that $\delta(X, S') \geq D$ for any length-n substring S' of S in $\mathcal{M}$?*

Let $\mathcal{M} = \{S_1, S_2, \dots, S_p\}$. We use $S[i]$ to denote the ith letter in string S. Let $\Sigma = \{a_1, a_2, \dots, a_m\}$. Let

$$\varphi[1] = \wedge_{\substack{1\leq i\leq p,\\ 1\leq j\leq |S_i|}} \vee_{1\leq k\leq m} x[i,j,k],$$

$$\varphi[2] = \wedge_{\substack{1\leq i\leq p,\\ 1\leq j\leq |S_i|,\\ 1\leq k[1]<k[2]\leq m}} (\neg x[i,j,k[1]] \vee \neg x[i,j,k[2]]),$$

$$\varphi[3] = \wedge_{\substack{1\leq i\leq p,\\ 1\leq j\leq |S_i|,\\ 1\leq k\leq m, S_i[j]=a_k}} x[i,j,k],$$

$$\varphi[4] = \wedge_{\substack{1\leq i\leq p,\\ 1\leq j\leq |S_i|,\\ 1\leq k\leq m, S_i[j]\neq a_k}} \neg x[i,j,k],$$

$$\psi[1] = \wedge_{1\leq j\leq n} \vee_{1\leq k\leq m} y[j,k],$$

$$\psi[2] = \wedge_{\substack{1\leq j\leq n,\\ 1\leq k[1]<k[2]\leq m}} (\neg y[j,k[1]] \vee \neg y[j,k[2]]),$$

$$\rho[1] = \wedge_{\substack{1\leq i\leq p,\\ 1\leq r\leq |S_i|-n+1,\\ 1\leq s\leq D}} \vee_{r\leq j\leq r+n-1} z[r,i,s,j],$$

$$\rho[2] = \wedge_{\substack{1\leq i\leq p,\\ 1\leq r\leq |S_i|-n+1\\ 1\leq s\leq D\\ r\leq j[1]<j[2]\leq r+n-1}} (\neg z[r,i,s,j[1]] \vee \neg z[r,i,s,j[2]]),$$

$$\rho[3] = \wedge_{\substack{1\leq i\leq p,\\ 1\leq r\leq |S_i|-n+1\\ 1\leq s\leq D\\ r\leq j\leq r+n-1,\\ 1\leq k\leq m}} (\neg z[r,i,s,j] \vee \neg x[i,j,k] \vee \neg y[j,k]),$$

$$\xi = (\wedge_{i=1}^{4}\varphi[i]) \wedge (\wedge_{i=1}^{2}\psi[i]) \wedge (\wedge_{i=1}^{3}\rho[i]).$$

It is easy to check that ξ is satisfiable if and only if there is a string X of length n over Σ such that $\delta(X, S') \geq D$ for any length-n substring S' of S in $\mathcal{M}$. Clearly, ξ is a CNF. So, ξ give us an explicit reduction from FSS to SAT.

Using standard transformations (see e.g. [2]) we can easily obtain an explicit transformation ξ into ζ such that $\xi \Leftrightarrow \zeta$ and ζ is a 3-CNF. It is clear that ζ gives us an explicit reduction from FSS to 3SAT. In papers [22, 23] the authors considered some satisfiability algorithms. Our computational experiments have shown that these algorithms can be used to solve FSS.

ACKNOWLEDGEMENTS. The work was partially supported by Analytical Departmental Program "Developing the scientific potential of high school" 8.1616.2011.

References

[1] A. Gorbenko and V. Popov, On the Longest Common Subsequence Problem, *Applied Mathematical Sciences*, 6 (2012), 5781-5787.

[2] A. Gorbenko and V. Popov, The Longest Common Parameterized Subsequence Problem, *Applied Mathematical Sciences*, 6 (2012), 2851-2855.

[3] A. Gorbenko and V. Popov, The set of parameterized k-covers problem, *Theoretical Computer Science*, 423 (2012), 19-24.

[4] V. Yu. Popov, Computational complexity of problems related to DNA sequencing by hybridization, *Doklady Mathematics*, 72 (2005), 642-644.

[5] V. Popov, The approximate period problem for DNA alphabet, *Theoretical Computer Science*, 304 (2003), 443-447.

[6] V. Popov, The Approximate Period Problem, *IAENG International Journal of Computer Science*, 36 (2009), 268-274.

[7] V. Popov, Multiple genome rearrangement by swaps and by element duplications, *Theoretical Computer Science*, 385 (2007), 115-126.

[8] A. Gorbenko and V. Popov, SAT Solvers for the Problem of Sensor Placement, *Advanced Studies in Theoretical Physics*, 6 (2012) 1235-1238.

[9] A. Gorbenko and V. Popov, On the Problem of Sensor Placement, *Advanced Studies in Theoretical Physics*, 6 (2012), 1117-1120.

[10] A. Gorbenko and V. Popov, The Problem of Selection of a Minimal Set of Visual Landmarks, *Applied Mathematical Sciences*, 6 (2012), 4729-4732.

[11] A. Gorbenko and V. Popov, On the Problem of Placement of Visual Landmarks, *Applied Mathematical Sciences*, 6 (2012), 689-696.

[12] A. Gorbenko and V. Popov, Self-Learning Algorithm for Visual Recognition and Object Categorization for Autonomous Mobile Robots, *Lecture Notes in Electrical Engineering*, 107 (2012), 1289-1295.

[13] A. Gorbenko and V. Popov, On Face Detection from Compressed Video Streams, *Applied Mathematical Sciences*, 6 (2012), 4763-4766.

[14] A. Gorbenko and V. Popov, Usage of the Laplace Transform as a Basic Algorithm of Railroad Tracks Recognition, *International Journal of Mathematical Analysis*, 6 (2012), 2413-2417.

[15] A. Gorbenko and V. Popov, Robot Self-Awareness: Usage of Co-training for Distance Functions for Sequences of Images, *Advanced Studies in Theoretical Physics*, 6 (2012) 1243-1246.

[16] A. Gorbenko and V. Popov, Robot Self-Awareness: Formulation of Hypotheses Based on the Discovered Regularities, *Applied Mathematical Sciences*, 6 (2012) 6583-6585.

[17] A. Gorbenko and V. Popov, Footstep Planning for Humanoid Robots, *Applied Mathematical Sciences*, 6 (2012) 6567-6571.

[18] A. Gorbenko and V. Popov, Multi-agent Path Planning, *Applied Mathematical Sciences*, 6 (2012) 6733-6737.

[19] A. Gorbenko and V. Popov, Programming for Modular Reconfigurable Robots, *Programming and Computer Software*, 38 (2012), 13-23.

[20] A. Gorbenko, M. Mornev, and V. Popov, Planning a Typical Working Day for Indoor Service Robots, *IAENG International Journal of Computer Science*, 38 (2011), 176-182.

[21] J. K. Lanctot, M. Li, B. Ma, S. Wang, and L. Zhang, Distinguishing string selection problems, *Information and Computation*, 185 (2003), 41-55.

[22] A. Gorbenko and V. Popov, Computational Experiments for the Problem of Selection of a Minimal Set of Visual Landmarks, *Applied Mathematical Sciences*, 6 (2012), 5775-5780.

[23] A. Gorbenko and V. Popov, Task-resource Scheduling Problem, *International Journal of Automation and Computing*, 9 (2012), 429-441.

Received: December 3, 2012